THE EQUALIZATION PROBABILITY OF THE PÓLYA URN

TIMOTHY C. WALLSTROM

ABSTRACT. We consider a Pólya urn, started with b black and w white balls, where $b > w$. We compute the probability that there are ever the same number of black and white balls in the urn, and show that it is twice the probability of getting no more than $w - 1$ heads in $b + w - 1$ tosses of a fair coin.

An urn contains b black and w white balls, where $b > w$. A ball is drawn from the urn at random, and then replaced with two balls of the same color. This same procedure is then repeated indefinitely. What is the probability that the urn ever contains the same number of black and white balls?

The problem involves the famous Pólya urn [3, 8, 10, 13]. The solution is not obvious, because the probability of a black or white draw is constantly changing, and an infinite number of different draw sequences can lead to equalization. The purpose of this note is to show that there is a remarkably simple solution: the equalization probability is just twice the probability that in $b + w - 1$ tosses of a fair coin, no more than $w - 1$ will be heads.

A probabilist would solve this problem by noting that the draws of the Pólya urn are *exchangeable*, in the sense that the probability of drawing any finite sequence of black and white balls depends only on the total number of black balls, and not on the order in which they are drawn. She would then invoke de Finetti's theorem, which states that an exchangeable process is a mixture of independent and identically distributed (i.i.d.) processes, which in this case means Bernoulli processes, or biased random walks [1, 5, 9]. In this way, she would reduce the equalization problem for the Pólya urn to the gambler's ruin problem, whose well-known solution dates to the inception of probability theory [4]. Such an approach is not without its charms, and also leads to efficient solutions of more difficult problems, such as the probability that there are ever k more white than black balls. However, it depends on the gambler's ruin results, and it seems worthwhile, if possible, to prove the result directly.

In this note, I provide an elementary proof of the main result, which does not depend on the gambler's ruin results. This work was inspired by a recent paper of Antal, Ben-Naim, and Krapivsky [2], who posed the equalization problem while working in the context of first-passage theory, and provided a closed form expression for the solution. The expressions in the current paper, however, are new.

Let $S_n = B_n - W_n$, the excess of black over white balls after n draws. S_n is a Markov process that starts at $S_0 = b - w > 0$, and at each step either increases by one, if a black ball is drawn, or decreases by one, if a white ball is drawn. The probabilities of these two events are just B_n/N_n and W_n/N_n, respectively, where $N_n = b + w + n$ is the number of balls in the urn after n draws. The trajectory of

2000 *Mathematics Subject Classification.* 60J80,60G09.
Key words and phrases. Pólya urn, exchangeability, first passage, beta distribution.

S_n thus resembles a random walk, except that the probabilities of moving up and down change with every step. We are interested in the probability that the path ever touches the S-axis.

We will need one key fact about the process S_n, or equivalently, about B_n: the fraction of black balls, B_n/N_n, has a random limit Z, which is given by the $\mathrm{Beta}_{b,w}$ distribution:

$$\mathrm{Beta}_{b,w}(p) = \frac{\Gamma(b+w)}{\Gamma(b)\Gamma(w)}\, p^{b-1}(1-p)^{w-1} \quad (0 < p < 1), \tag{1}$$

An elementary proof may be obtained by following the reasoning in de Finetti [6, p. 219], who shows how the Pólya urn process may be modeled using random draws from the uniform distribution. For a less elementary proof, see [7, §2]. It follows that $\mu_n = S_n/N_n$ also converges, to $\mu \equiv 2Z - 1$. We write $F^{\mathrm{Beta}}_{b,w}(p) = \int_0^p \mathrm{Beta}_{b,w}(p)\, dp$ for the distribution function of the beta distribution.

Theorem. *If a Pólya urn is started with b black and w white balls, then the probability that the number of black and white balls will ever be equal is $2F^{Beta}_{b,w}(\frac{1}{2})$.*

Proof. Let τ be the random time at which the path first touches the boundary $S = 0$. The probability of equalization is just the probability of the event $\{\tau < \infty\}$, which can be divided into the two events $\{\tau < \infty\} \cap \{\mu > 0\}$ and $\{\tau < \infty\} \cap \{\mu < 0\}$. (The probability that $\mu = 0$ is zero, because the density in Eq. (1) is continuous.) But these two events have equal probability. Indeed, μ depends only on $S_\tau(n) \equiv S(\tau + n)$, because the initial segment is finite, and has no effect on the mean. At time τ, the urn has an equal number of black and white balls, so the mean of its subsequent trajectory is equally likely to be positive or negative. Furthermore, if $\mu < 0$, then $\tau < \infty$. Indeed, if $\mu < 0$, then the path will eventually be below the axis, and since it started out above the axis, it must cross at some point. Thus,

$$P(\tau < \infty) = 2P(\{\tau < \infty\} \cap \{\mu < 0\}) = 2P(\{\mu < 0\}) = 2F^{\mathrm{Beta}}_{b,w}(\tfrac{1}{2}).$$

The last expression follows from the fact that $\mu < 0$ if and only if $p < \frac{1}{2}$. $\square$

The equalization probability can also be expressed as a binomial sum, due to an interesting connection between the beta and uniform distributions. Let $U_1, U_2 \ldots, U_n$ be n independent samples from the uniform distribution on $[0, 1]$, and let $U_{(1)} < U_{(2)} < \cdots < U_{(n)}$ be the same samples arranged in increasing order. (The $U_{(i)}$ are called the *order statistics* of the sample.) Then the density of $U_{(b)}$ is given by $\mathrm{Beta}_{b,w}$, where $b + w = n + 1$. We refer the reader to [5, Eq. I.6.7] for the simple proof.

Given this result, the probability that a $\mathrm{Beta}_{b,w}$ variable will be less than x is just the probability that at least b uniform variates will be less than x, or equivalently, that no more than $n - b$ uniform variates will be greater than x. Let Ber_p denote a Bernoulli random variable, taking the value one with probability p, and zero otherwise. In symbols, then

$$P(\mathrm{Beta}_{b,w} \le x) = P(\sum_{i=0}^{b+w-1} X_i \ge b) = P(\sum_{i=0}^{b+w-1} Y_i \le w - 1), \tag{2}$$

where the X_i and Y_i are independent Ber_x and Ber_{1-x} random variables, respectively. The last expression, with $x = 1/2$, establishes the following corollary:

Corollary. *If a Pólya urn is started with b black and w white balls, then the probability that the number of black and white balls will ever be equal is the same as the probability that in $b+w-1$ tosses of a fair coin, no more than $w-1$ will be heads.*

Eq. (2) was first derived computationally by Karl Pearson in 1924 [12], for the purpose of expressing sums of binomial coefficients in terms of the more easily computable beta distribution. See also [4, p. 173], [11, 8.17.5].

Both expressions for the equalization probability are useful. The first, involving the beta function, is easily computed numerically. The second can be used to establish central limit results, using the deMoivre-Laplace theorem [4], or large deviations results, using Cramér's theorem [9]. This second expression can also be written as an explicit sum of binomial coefficients,

$$\frac{1}{2^{b+w-2}} \sum_{j=0}^{w-1} \binom{b+w-1}{j} = 1 - \frac{1}{2^{b+w-1}} \sum_{j=w}^{b-1} \binom{b+w-1}{j},$$

and these forms are useful when either w or $b-w$ are small, respectively.

Acknowledgments

I thank Eli Ben-Naim for helpful comments, and acknowledge support from the Department of Energy under contract DE-AC52-06NA25396.

References

[1] David J Aldous, *Exchangeability and Related Topics*, Ecole d'Ete de Probabilites de Saint-Flour XIII, 1983 (Lecture Notes in Mathematics 1117), Springer, 1985.
[2] T Antal, E Ben-Naim, and PL Krapivsky, *First-passage properties of the Pólya urn process*, Journal of Statistical Mechanics: Theory and Experiment **2010** (2010), P07009.
[3] F Eggenberger and G Pólya, *Über die Statistik verketteter Vorgänge*, ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik **3** (1923), no. 4, 279–289.
[4] William Feller, *An Introduction to Probability Theory and Its Applications, Vol. 1*, 3rd ed., Wiley, 1968.
[5] ———, *An Introduction to Probability Theory and Its Applications, Vol. 2*, 2nd ed., Wiley, 1971.
[6] Bruno De Finetti, *Theory of probability. a critical introductory treatment. Volume 2.*, 1st ed., John Wiley & Sons Ltd, 1975.
[7] DA Freedman, *Bernard Friedman's urn*, The Annals of Mathematical Statistics **36** (1965), no. 3, 956–970.
[8] Norman L. Johnson and Samuel Kotz, *Urn models and their application: An approach to modern discrete probability theory*, John Wiley & Sons Inc, 1977.
[9] Olav Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, 2010.
[10] Hosam Mahmoud, *Polya Urn Models*, 1st ed., Chapman and Hall/CRC, 2008.
[11] Frank W J Olver, Daniel W Lozier, Ronald F Boisvert, and Charles W Clark, *NIST Handbook of Mathematical Functions*, Cambridge University Press, 2010.
[12] K Pearson, *Note on the Relationship of the Incomplete B-function to the Sum of the first p terms of the Binomial* $(a+b)^n$, Biometrika (1924jan), 202–203.
[13] G Pólya, *Sur quelques points de la théorie des probabilités*, Annales de l'Institut Henri Poincaré **1** (1930), no. 2, 117–161.

Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545
E-mail address: tcw@lanl.gov